



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

"TEACHERS' MARKS."

By C. S. BRAGDON.

The chairman of the program committee suggested as an appropriate topic for this occasion a review of some recent book or books on educational topics of interest to mathematics teachers. As I was browsing among the newer books on education in our library a little volume attracted my attention. It was entitled "Teachers' Marks"—by Frederick James Kelly, Ph.D. Most teachers consider the marking of papers a bore and the least interesting phase of a teacher's work. What could be more tedious, therefore, than a scientific treatise, composed of tables of figures and per cents. and charts on this subject? But just as nuggets of gold are sometimes found in unexpected places so this seemingly unattractive book was found to contain material of unusual interest. Because it interested me, I concluded that it might not prove unattractive to you teachers of mathematics who are supposed to revel in figures and statistics.

I shall not attempt a review of the book, but allow it to supply us certain facts and figures which will form the basis of a brief discussion of the topic suggested in the title.

Passing over the first chapter on grading the work in elementary schools to the second which treats of standards of marking in high schools, Mr. Kelly quotes from a study made by Mr. F. W. Johnson of the University High School of Chicago figures to show the wide divergence of marks given by two teachers in the same subject in that school. The first teacher had 8 per cent. failures and 7.5 per cent. A's or highest grade; the second had 4.5 per cent. failures and 36 per cent. A's. When teachers in different subjects were compared a still greater divergence in marks was found; one teacher having 26.5 per cent. failures and 1.5 per cent. A's as compared with 4.5 per cent. failures and 36 per cent. A's for another teacher.

That the passing standard is largely a matter of tradition or whim is the conclusion the author draws from the statistics of a

certain New York City high school. A change of principals was made in this school in 1910. In the previous year only 48 per cent. of the algebra pupils were passed the first term and 61 per cent. the second term, while the new principal decided to pass 75 per cent. of the algebra pupils the first term and 80 per cent. the second.

A series of tables then follows showing the successive marks received by pupils from grade to grade, through high school and college. The discovery is made that only about 50 per cent. of the large number of cases studied retain the same relative position by thirds or tertiles, *i. e.*, of those in the upper third of the elementary grades only about one half maintained a place in the upper third of their classes in high school, etc. From this study the author draws the following conclusion: "If we can come no nearer than that in ranking our children for general ability, we cannot hope to command much respect as a teaching profession. Rather should the revelations made by these studies open our eyes to the real need for some more effectual method for establishing standards whereby both teachers and pupils may measure progress."

Passing next to the marking of examination papers, he says: "The few studies which have been made reveal a very wide difference of rating upon the same papers among supposedly competent judges." In support of this contention he cites the experiment made by an Oxford professor who caused to be inserted in the *English Journal of Education* a specimen of Latin prose composition. He then invited competent judges to rate the paper and send him their results. Twenty-eight replies were received with the marks as follows: 45, 59, 67, 67.5, 70, 70, 72.5, 6 at 75, 77, 5 at 80, 2 at 82, 2 at 85, 87.5, 88, 90, 2 at 100.

The results of an experiment at Teachers' College, Columbia University, also confirm the above statement. Eleven judges, all graduate students, were asked to rate a set of twenty papers in geography. The marks on two of these papers will be sufficient to illustrate the diversity of ratings. Paper No. 4 was rated by the eleven judges as follows: 27, 50, 60, 28, 59, 60, 48, 40, 90, 72, 15, 50. Paper No. 17—53, 50, 90, 54, 93, 63, 46, 60, 100, 39, 100, 59.

At the University of Wisconsin the following experiment was

tried: A facsimile reproduction was made of a geometry paper written by a pupil in one of the leading high schools of the state. Copies were then sent to a large number of schools with the request that the paper be rated by the teacher in each school best qualified for the work. The 116 replies contained marks as follows: 1 at 28, 1 at 39, 1 at 41, 1 at 44, 2 at 48, 6 at 50, 6 at 54, 8 at 55, 8 at 59, 17 at 60, 17 at 64, 19 at 65, 19 at 69, 7 at 85, 7 at 89, 2 at 90, 2 at 94.

I will quote but one other experiment in marking of a set of papers in mathematics from an Indiana high school. The regular teacher had given to this set an average mark of 78.7 per cent. Five other competent persons were asked to rate the same papers. One gave to the set an average mark of 74 per cent., the second, 61.4 per cent., the third, 65.5 per cent., the fourth, 75.5 per cent., the fifth, 58.0 per cent.

Mr. Kelly then says: "In all of the above studies we see very serious lack of standards among teachers. It is true that in all these cases the judges were selected from an area where no especial effort had been made to standardize the judgments. On this account, I undertook to measure the variations between the marks of teachers in New York State on the one hand and of the regents on the other." He then quotes statistics from the regents' reports for several years up to 1913 and says: "The two tendencies to which attention is called are the constantly increasing per cent. of papers which the regents have passed and, at the same time, the constantly increasing per cent. of papers rejected by the regents of those passed by the teachers. These two tendencies seem to me significant. While an ever increasing number of pupils in the high schools of the state are able to meet the requirements of the examiners, the difference in the standards of judging papers by teachers and examiners grows ever greater. While the requirements for high-school teachers are constantly being increased, their judgment of the value of examination papers is being more and more rejected. The greater the per cent. failed by the teachers, the greater the additional per cent. failed by the regents. The rule is not even violated in the case of mathematics which by all tradition offers the greatest possibility of exactness in marking papers." Continuing, he says: "It is a certain indication that there is as little

agreement among the teachers of the state concerning standards hoped for by the regents themselves in the examinations as there is among the examiners in the various subjects."

He then asks a question concerning the fate of papers rated by the teachers around 60 per cent. and finds these figures to be true: "41.3 per cent. of such papers were failed by the regents, 5.64 per cent. raised above 65 per cent. and more than one half left at 60 per cent." "The chief interest," says he, "in this table lies in the report common among high-school teachers that they push up the grade on doubtful papers to take a chance on their passing. Results seem to indicate that the policy is a wise one, for of papers marked at 60 per cent. the lowest per cent. saved to any school was 22 per cent., the highest 75 per cent., the average 58.7 per cent. Finally, by all these findings concerning the New York state system of examinations we are compelled to conclude that the type of examination now in common use is *not* a successful means of standardizing school achievement."

The writer then proceeds to suggest a better means of standardizing the results of examinations, as shown by an experiment in the schools of Orange, N. J. A uniform test was given to all fifth-grade pupils in arithmetic. These papers were rated by the respective teachers. Afterwards a most efficient and sympathetic teacher was asked to rate all the papers and compare her results with the results of the others. Afterwards this teacher was asked to submit an appropriate scheme for marking each question on the test. Then the first teachers were asked to re-rate their papers using this scheme. The results showed a very considerable range of difference when teachers used their own standards of marking, but when the same scheme is used by both teachers and judge the range of differences is very much reduced, considerably more than half the cases being zero. "From this experiment," he says, "we may draw one lesson: If the superintendent expects to place much significance upon the uniform tests he gives he must either have the marking done by a single judge, or else he must make out a scale for the rating of the papers by which the variations of the several teachers may be greatly reduced."

In the remaining fifty pages the author examines the various

devices for standardizing results, the most interesting ones to teachers of mathematics being those of Stone and Courtis pertaining to arithmetic. No discussion of them, however, will be attempted in this paper, as I believe it will be more profitable to spend the remainder of our time in considering some of the statements and criticisms already quoted.

We may summarize the conclusions of Mr. Kelly under the following heads:

I. There is wide variation in rating work of a definite grade, due to the point of view or personal equation of the teacher.

II. Such wide variation constitutes a serious reflection on the standing of teaching as a profession, since those supposed to be experts vary in marking the same paper between 40 per cent. and 100 per cent.

III. The examination as a means of measuring the proficiency of pupils and furnishing a basis of promotion is of little value unless a method of standardizing results may be devised.

IV. The regents system of uniform examinations is failing to provide the desired standardization as shown by the great differences between the judgment of the teachers and that of the examiners of the state department.

The first three conclusions will readily be granted. The fourth, concerning the failure of the regents system to secure uniformly reliable results, contains much truth. The author, however, neglects to state the fact that the examinations division urges a system of marking which, if generally followed, would assist greatly in securing the desired results. I refer to the committee system of marking with which you are undoubtedly familiar. To be sure, more time is required to rate a set of papers by this method, but the results are enough better to warrant the additional time and effort. It does seem, however, that after three competent teachers have agreed upon a scale of marking and then have carefully rated a set of papers by the committee system that their combined judgment of the marks to be given must be as nearly accurate as it is possible to obtain. Why, then, the need of having such papers re-rated at Albany? Does it not mean the setting aside of the judgment of a committee of competent examiners and substituting therefor the judgment of an individual no more competent than any one of the

committee? Do not the statistics quoted from Mr. Kelly's book prove that the judgment of an individual, no matter how competent, is less likely to be correct than the combined judgment of several? Would it not therefore be safe for the examiners to accept at face value, without re-examination, all papers which have been carefully marked by the committee system? Would there not be a very strong inducement for more schools to adopt the committee scheme of marking if they knew that their marks would be accepted. If the combined judgment of three teachers claims a paper at 64 and then a single examiner rejects it at 58 what inducement is there to schools to mark by the committee system?

Not long ago I received a letter from a high-school principal in which he said: "We would like to rate our papers about as they will be rated at Albany, but find it quite impossible. Are you having the same experience?" It seems to me that this little book by Mr. Kelly proves that agreement between the schools and the examinations division is impossible so long as the mark of an individual teacher in any school is compared with that of an individual examiner at Albany. This statement is in no sense a reflection on the judgment of the teacher nor on that of the examiner. It is simply the inevitable consequence of the difference in points of view of two equally competent judges working independently without a prearranged standard of marking. Failure to appreciate the full significance of this great truth frequently leads to mutual distrust of the examiners by the schools and of the schools by the examiners. Such a condition ought not to exist and would not if a closer understanding between teachers and examiners could be effected. The sentiments expressed by the high-school principal quoted above are shared by all principals. We would all like to have our teachers rate their papers as the examiners wish them to be rated. And right here I wish to protest against a current notion that we are concerned merely with the number of papers accepted or rejected. We are just as anxious, yes, even more anxious to know why a paper rated by our teachers at 93 is reported back with a mark of 81 as to know why a paper just on the border line of failure is rejected. Especially is this true since such a reduction of ten or more points may mean for a pupil the loss of a state

scholarship. The privilege of reviewing high papers thus reduced in mark ought to be used much more than is now the case, because by so doing a better mutual understanding of systems of marking may be realized. The real purpose of the examinations is not to determine how many pass or fail, but to give a just and reasonably accurate rating to all papers, high or low. And teachers and examiners must come together on this matter and agree on standards understood by all so that such differences in rating may be avoided.

But some may say that such standards have been established by the specialists in the various subjects. Quite recently a new series of pamphlets entitled "Suggestions on the Rating of Regents Examinations" has been forwarded to the schools. We all hoped that these new suggestions would be sufficiently explicit and specific to constitute a standard for rating papers, and thus eliminate largely the possibility of widely divergent marks for papers of equal merit. But a careful reading of these suggestions shows them to be disappointing in this respect. In the suggestions for rating mathematics papers several statements are made which are capable of widely differing interpretations, so that we are just as far from a definite standard by which to rate our mathematics papers as we have been in the past. Until such a definite standard has been adopted our uniform examinations can never bring uniform justice to our schools nor to the pupils in our schools.

That the examinations division appreciate the truth of this assertion is evidenced from the following statement from Mr. Horner, chief of the examinations division, in his report for 1913. He says: "Something more than suggestions is needed to secure a fair degree of uniformity in rating answers." He then advocates the use of the committee system to which I have referred as a means of preventing serious differences in rating that are due to differences in temperament. In the same report Mr. Horner says: "It is hardly too much to hope that registered secondary schools may some time be brought to such a high standard that the local ratings will be final in all cases." Surely the examinations division is not more anxious to see this hope realized than are the principals and teachers of our secondary schools, yet according to the 1914 report only 57 schools out of nearly 900 had 95 per cent. or more of its papers claimed ac-

cepted. In other words, less than 7 per cent. of the secondary schools of the state seemed to have a sufficiently clear comprehension of the standard of the examinations division to have 95 per cent. of its claimed papers accepted, while in 150 secondary schools not more than 70 per cent. and in some schools as low as 45 per cent. of its papers claimed were accepted.

These figures force us again to the conclusion that some means should be found for more clearly defining the standard of the examinations division, for if that standard were clearly understood at least 50 per cent. of the schools of the state ought to be able to rate their papers by the committee system in harmony with the standard. What a saving of time and money this would mean to the state for, to quote Mr. Horner again, "It is worse than useless for the department examiners to spend their time in rereading thousands of papers rated by teachers fully as competent and as accurate in rating as they are."

In closing may I venture to suggest another possible means of securing greater uniformity in rating regents' papers. The examinations in mathematics, for example, are compiled by a committee of experienced teachers acting with the specialist in mathematics of the department. What body could be more competent than this to formulate a definite scale for marking the various parts or questions on these papers? Specific directions for marking each mathematics examination could thus be formulated by this committee; these directions could be printed and a few copies sent to each school with the question papers. Would not such a set of specific directions for marking a particular examination tend to secure far greater uniformity in marks than a set of general suggestions intended for several different types of examinations? Possibly some teachers might object to such a plan on the plea that it would deprive them of their individuality and make them mere marking machines. But if uniformity in marking is desirable, then the personality of the marker must become of less relative importance than conformity to the standard. For the marks should show the true relative worth of the work of an individual pupil in comparison with the work of all others taking the same examination. In this way only can justice be secured.

UTICA FREE ACADEMY,
UTICA, N. Y.